VALIDATION OF A LATENT CLASS ANALYSIS MODEL FOR PREDICTING MORTALITY[1]

DONALD B. WITZKE, LUIS CASTRO AND HARVEY F. DINGMAN
University of Texas at Austin

Analysis of data pertaining to mortality in hospital populations has been a persisting problem. Generally, it has been demonstrated that the relevant variables involved in the classification of differential mortality are sex, race, age, IQ, and diagnosis. When cross-classifications have been made on these variables holding age constant, however, the small sample sizes which resulted usually restricted further classification with reference to mortality to gross estimates.

In order to analyze the joint effect of sex, age, race, IQ, and diagnosis on differential mortality, Miller, Sabagh, and Dingman (1962) have suggested the use of latent class analysis. The rationale for using this method was based on Lazarsfeld's original assumptions that "observed" attributes are assumed to reflect underlying attributes which have point distributions, and that individuals who are identified at any given point may be said to belong to a latent class. One of the basic assumptions is that their correlations are a function of the varying response probabilities in the different latent classes. Therefore, the people who have similar response patterns are assigned to a latent class on the basis that they probably have identical response patterns.

This approach was used by Miller, *et al.* (1962) in identifying differential mortality in a mentally retarded population. Two groups were identified, based on their latent class structure, determined by the joint occurrence of sex, age, race, IQ, and diagnosis. As a result of their mathematical solution for the latent class model, a set of weights (based on the mentally retarded population) were generated. These weights, which are displayed in Table 1, represent the proportions of patients in each latent class theoretically expected to have specified characteristics.

Because of the theoretical and practical importance implicit in the ability to use a typological scheme to identify probable mortality in hospital patients, latent class analysis may significantly serve to

[1]Supported in part by the Texas Foundation for Mental Health Research in collaboration with the Texas Department for Mental Health and Mental Retardation. Mental Health Project Grant Number 09235-01 of the National Institute of Mental Health. The authors gratefully acknowledge Dr. Alex Pokorny and the Administrative Survey of Texas State Mental Hospitals for the use of patient data.

Table 1
Theoretical weights for patient response characteristics for each latent class

Patient Characteristic Response	Latent Class			
	I		II	
	Yes	No	Yes	No
Sex (Female)	.400	.600	.477	.553
Age (Under 10 years old)	.153	.874	.723	.277
Race (White)	.650	.350	.933	.067
IQ (Under 40)	.255	.745	.924	.076
Diagnosis (Somatic)	.277	.723	.892	.108

augment clinical prognostications (O'Connor, G., & Dingman, H. F., 1966). Further, the importance of mortality suggests that cross-validation must be made of the procedures, as well as the equations, used in the identification of differential mortality on the basis of latent class analysis.

The purpose of the present study is to test the generality of the proposed model in the identification of latent class structure in relation to differential mortality by using the weights obtained on a mentally retarded population and applying these weights to a mentally ill population.

METHOD

The subjects used in this study were selected from a larger sample of 1537 patients studied in the Administrative Survey of Texas State Mental Hospitals in 1966. After one year had passed patient movement records were compiled, and from these records all patients with mortality clasisfications were selected. After those patients who had missing information on any criterion variable were deleted from the study, a mortality group with an N of 100 resulted. An additional 100 non-mortality patients were drawn at random from the larger sample. All patients had a primary classification of mental illness. Each of the criterion variables was dichotomized as follows: (1) female and male, (2) below 40 years old and above 39 years old, (3) white and other, (4) below dull normal IQ and above borderline IQ, and (5) somatic and non-somatic diagnosis.[2] Table 2 shows

[2]IQ was determined by a panel of psychologists who rated each patient on the basis of his responses to the Logical Memory (Story A of Wechsler Memory, Form I), Kent E-G-Y (Scale D), Bender Gestalt, Otis Arithmetical Reasoning Test (Form A), Benton Visual Retention Test (Form E), and Human Figure

the frequency for each these dichotomies by mortality and non-mortality group.[3]

Table 2

Frequencies for each of the five variables for the 100 persons in Texas State Mental Hospitals that did not die in the critical period, and for the 100 residents in Texas State Mental Hospitals that *did* die during the time period under examination.

Patient Characteristic	Mortality N = 100	Non-mortality N = 100	Total
Female	46	55	101
Male	54	45	99
Below Age 40	5	25	30
Above age 39	95	75	170
White	76	75	151
Other	24	25	49
Below dull normal IQ	82	56	138
Above borderline IQ	18	44	62
Somatic	67	17	84
Non-somatic	33	83	116

After the sample selection was made, each patient was assigned to a latent class according to the weights presented in Table 1 based on the response made to each of the criterion variables (or patient characteristics). The assignment of each patient to a latent class was made using the computational formulas which follow.

Drawings. The remaining variables were established according to the criteria used by Miller, *et al.* (1962).

[3]Because the formula was somewhat inappropriate as it was based on a population of mentally retarded that was substantially younger and that had lower IQ's, the dichotomies for age and IQ were established at a somewhat different point for the present study.

i = item

s = class

P_s = joint probability for assignment to a given class

II_s = probability of belonging in a given class

V_{is} = v_{is} if the response to an item is positive

V_{is} = $(1-v_{is})$ if the response to an item is negative

$$P_s = V_{1s} \cdot V_{2s} \cdot V_{3s} \cdot V_{4s} \cdot V_{5s}$$

$$II_1 = P_1 / P_1 + P_2$$

$$II_2 = P_2 / P_1 + P_2$$

RESULTS

After determining the probability that a person belonged to either Class 1 or to Class 2, a four-fold table (Table 3) was derived and a Chi Square was computed to determine if the cells differed significantly from a random distribution. The results of the Chi square

Table 3

Assignment to latent classes of the persons that died in Texas State Mental Hospitals during the critical period as well as those who who did not die during the period under examination.

	Mortality	Non-mortality	Total in each class
Class 1	54	83	137
Class 2	46	17	63
Total	100	100	

Chi Square = 19.4879, df = 1, $p < .0001$

indicated that the obtained frequencies were significantly different from an expected random distribution. Although the classifications for the mortality group were not appreciably different between classes, only 17% of the patients in the non-mortality group were misclassified as belonging to Latent Class 2. The 83% of the non-mortality patients who were correctly identified as belonging to

Latent Class 1 tended to be those who were males, below age 40, not white, above borderline IQ, and diagnosed non-somatic.

DISCUSSION

The ability to identify, by latent class membership, those patients who have similar characteristics in relation to mortality suggests that the proposed latent class model can be generalized. Because the weights which were used to classify mentally ill patients were developed on a mentally retarded population, a more accurate solution to the latent class structure might be found by generating weights for a mentally ill population. In addition, there may also be more critical points for dichotomizing each of the criterion variables; or further, there may be other variables which are more germane to a mentally ill population with reference to the problem of latent class structure. These suggestions allude to the possibility that the placement of mortality and non-mortality patients into improper latent classes may be in a part a function of these problems.

When mortality rate is related to length of stay and latent class. there is a tendency for a greater number of deaths to occur within the first five years of stay in the hospital. Although the differentia-

Table 4

Mortality and non-mortality rate for each latent class by length of stay.

Length of Stay	Class I		Class II		Total for each time period
	Mortality	Non-mortality	Mortality	Non-mortality	
90 days or less	9	19	7	3	38
91 to 365 days	8	5	4	2	19
1 to 5 years	17	17	21	3	58
6 to 10 years	7	7	10	2	26
11 to 20 years	6	20	2	6	34
21 to 30 years	2	14	0	1	17
31 to 40 years	2	1	1	0	4
41 to 50 years	3	0	1	0	4

tion by latent class for the mortality group, as shown in Table 4, is not extreme, there are marked differences for the comparisons between the mortality and non-mortality groups. The non-mortality groups tend to have a longer length of stay.

In general, the findings reported in the present study tend to

support the suggestion that the latent class model for identification of differential mortality is appropriate for different hospital populations. It is further suggested, however, that a better solution to the problem of latent class assignment may be made if both the criterion variables and their relative weights are established on the populations for which they are to be used.

REFERENCES

Miller, C. R., Sabagh, G., & Dingman, H. F. Latent class analysis and differential mortality. *Journal of the American Staistical Association*, 1962, 57, 430-438.

O'Connor, G., & Dingman, H. F. Statistical typologies as a supplement to clinical diagnosis. *Behavioral Science*, 1966, 11, 342-345.

ABSTRACT

An equation to predict mortality in institutions has been developed by Miller, Sabagh, and Dingman (1962). This equation was cross validated on a population of mentally ill in a state hospital in another state. The equation provided successful predictions of mortality ($p < .0001$) even though the weights had been derived for a younger and more severely handicapped group. It was concluded that a better understanding and better predictions would follow if the analysis had been based on the mentally ill population.

RESUMEN

Miller, Sabagh y Dingman (1962) han desarrollado una educación para poder pronosticar la mortandad en instituciones. Esta educación fue validada entre una población de enfermos mentales en un hospital estatal en otro estado. La educación dio buenos pronósticos de mortandad ($p < .0001$) aun cuando se habían obtenido las pesas para un grupo más joven y con defectos mentales más severos. Se llegó a la conclusión que resultaría un mejor entendimiento y un mejor prognóstico si el análisis se hubiera basado en la población de enfermos mentales.

RESUMO

Uma equação para predizer em instituções hospitalares foi desenvolvida por Miller, Sabagh e Dingman (1962). Esta equação foi validada com uma população de pacientes mentais num hospital estadual. A equação demonstrou predições corretas de mortalidade ($p < .0001$) apesar do fato de que os pêsos haviam sido derivados para um grupo mais jovem e mais severamente debilitado. Concluiu-se que resultariam melhor entendimento e melhores predições se a análise tivesse sido baseada na população de pacientes mentais.